Solution structure of the porcine sapovirus VPg core reveals a stable three-helical bundle with a conserved surface patch

Hyo-Jeong Hwang\textsuperscript{a,1}, Hye Jung Min\textsuperscript{a,1}, Hyosuk Yun\textsuperscript{a}, Jeffery G. Pelton\textsuperscript{b}, David E. Wemmer\textsuperscript{b}, Kyoung-Oh Cho\textsuperscript{c}, Jeong-Sun Kim\textsuperscript{a,1}, Chul Won Lee\textsuperscript{a,1,*}

\textsuperscript{a} Department of Chemistry, Chonnam National University, Gwangju 500-757, Republic of Korea
\textsuperscript{b} Division of Physical Biosciences of Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720, USA
\textsuperscript{c} Laboratory of Veterinary Pathology, College of Veterinary Medicine, Chonnam National University, Gwangju 500-757, Republic of Korea

**A R T I C L E   I N F O**

Article history:
Received 16 February 2015
Available online 6 March 2015

Keywords:
NMR spectroscopy
Porcine sapovirus
RNA-dependent RNA polymerase
Viral protein genome-linked

**A B S T R A C T**

Viral protein genome-linked (VPg) proteins play a critical role in the life cycle of vertebrate and plant positive-sense RNA viruses by acting as a protein primer for genome replication and as a protein cap for translation initiation. Here we report the solution structure of the porcine sapovirus VPg core (VPg\textsubscript{c}) determined by multi-dimensional NMR spectroscopy. The structure of VPg\textsubscript{c} is composed of three \(\alpha\)-helices stabilized by several conserved hydrophobic residues that form a helical bundle core similar to that of feline calicivirus VPg. The putative nucleotide acceptor Tyr956 within the first helix of the core is completely exposed to solvent accessible surfaces to facilitate nucleotidylation by viral RNA polymerase. Comparison of VPg structures suggests that the surface for nucleotidylation is highly conserved among the Caliciviridae family, whereas the backbone core structures are different. These structural features suggest that caliciviruses share common mechanisms of VPg-dependent viral replication and translation.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

The Caliciviridae family is a member of single-stranded and positive-sense (+) RNA viruses and can be divided into at least four genera based on genome organization and genetic analysis: Vesivirus, Lagovirus, Norovirus, and Sapovirus [1,2]. Caliciviruses in a number of organisms cause a wide range of diseases in vertebrates. For instance, human norovirus (HuNV) causes acute gastroenteritis in humans [3], while the feline calicivirus (FCV) is related to upper respiratory infections in cats [4]. In particular, sapoviruses (SaVs) are a major cause of gastroenteritis worldwide in both humans and animals [5,6]. Although SaV infections have raised public health concerns of potential cross-species transmission [7], studies on the molecular mechanisms of viral infection and replication have been hindered by the lack of a cell culture system for human SaV. Therefore, a cultivable porcine sapovirus (PSV) system has been used to study the viral infectious cycle and molecular mechanisms of other SaVs including human SaV [8].

Viral protein genome-linked (VPg) proteins have been identified as a protein primer during RNA synthesis in many virus families including Caliciviridae and Picornaviridae [9,10]. VPg proteins are covalently linked to the 5\textsuperscript{\prime}-end of the viral genome or sub-genome of (+) RNA viruses by a phosphodiester bond between the hydroxyl group of a serine or tyrosine residue in VPg and the 5\textsuperscript{\prime}-end of uridine or guanine of the viral RNA [11,12]. Studies of the VPg proteins from various plant and vertebrate (+) RNA viruses revealed their critical roles in viral replication and translation [13,14]. In caliciviruses, the viral RNA polymerase can nucleotidylate the VPg protein [15], and the nucleotidylated VPg is extended to produce RNA-linked VPg in a template-dependent or -independent manner [16,17], indicating that caliciviral VPgs act as a protein primer for the genome synthesis of (+) RNA viruses. The nucleotide is linked to a conserved tyrosine residue: Tyr24, Tyr26 or/and Tyr117, Tyr27, and Tyr21 in the VPg of FCV, murine norovirus (MNV), HuNV, and rabbit hemorrhage disease virus (RHDV), respectively [15–19].

VPg proteins of caliciviruses are also involved in the initial stage of virus infection and viral protein synthesis. The (+) RNA viruses are immediately translated upon infection with the viral genome acting as an mRNA template. In this stage, the VPg covalently linked

---

\* Corresponding authors. Fax: +82 62 530 3389.
E-mail addresses: jsunkim@chonnam.ac.kr (J.-S. Kim), cwlee@jnu.ac.kr (C.W. Lee).

1 These authors contributed equally to this work.
to the viral genome binds to host translation initiation factors (IFs) and functions as a proteinaceous cap substitute. The VPg proteins from FCV and MNV can directly interact with the cap-binding protein elf4E [20,21]. The FCV VPg-elf4E interaction is essential for FCV translation, but the interaction between MNV VPg and elf4E is dispensable for viral translation initiation [20]. We recently demonstrated that the PSV VPg is essential for viral translation and infectivity by directly forming the VPg-elf4E complex [22], whereas the interaction between VPg and elf4G is required for MNV translation initiation [23]. These observations suggested that the caliciviral VPg proteins have diverse and complex roles in viral translation initiation and infection.

VPg proteins vary in size and sequence among viruses, indicating large differences in their respective structures. For example, the calicivirus VPgs are 13–15 kDa proteins that form a stable helical core structure [24]. On the other hand, the picornavirus VPgs are short peptides of 22–24 amino acids that have no distinct structure and are ordered upon interaction with RNA genome [14]. Recently, the structures of two VPg proteins from FCV and MNV that are closely related to PSV were determined using solution NMR spectroscopy [24]. Although the structured cores of the FCV and MNV VPgs are distinct, especially in terms of sequence length, they have a common compact helical core flanked by unstructured N- and C-terminal regions. The nucleotidylation Tyr residues are exposed to solvent and are located in a conserved position in the first helix of the core.

Our previous study clearly showed that the PSV VPg-elf4E interaction is required for PSV translation, indicating that PSV is functionally similar to FCV [22]. However, structures of PSV VPgs have not been reported, which hinders understanding of the molecular mechanism of PSV VPg-mediated viral infection and proliferation. In this study, we determined the solution structure of the PSV VPg core (VPgC).

2. Materials and methods

2.1. Cloning, expression, and purification of VPg

The genes encoding PSV VPgs (residues Ala935–Glu1048, Ala935–Gly1022, Ala945–Gly1022, and Ala948–Ser1006) were amplified from the PSV genome by polymerase chain reaction (PCR) (Fig. 1A). The PCR product was cloned into pPROEX-HTc (Invitrogen, Life Technologies). These expression constructs express 25 extra amino acids at the N-terminus including six continuous His residues and the Tobacco etch virus (TEV) cleavage sequence. The recombinant plasmids were transformed into the Escherichia coli B834(DE3) strain, which was grown in Luria–Bertani medium containing 100 μg mL⁻¹ ampicillin at 310 K. After induction by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside, the culture medium was maintained for a further 8 h at 310 K. Cells were harvested by centrifugation and re-suspended in buffer A (20 mM Tris–HCl at pH 7.5 and 500 mM...
NaCl), then disrupted by ultrasonication. The fusion protein was purified using a 5-mL HisTrap HP chelating column (GE Healthcare). After treatment with recombinant TEV protease to cleave the six His residue tag and removal of salts by dialysis, the protein, which contains three additional amino acids (Gly-Ala-Met) at the N-terminus, was purified using a 5-mL HiTrapQ anion-exchange column (GE Healthcare). The column was washed with buffer B (20 mM Tris–HCl at pH 7.5), and the bound protein was then eluted with a linear gradient of 0–1000 mM NaCl in buffer B. For the NMR experiments, 15N-labeled or 13C- and 15N-labeled proteins were expressed in M9 minimal medium and purified by applying the procedures described above.

### 2.2. NMR spectroscopy

For NMR experiments, the purified VPG proteins were dissolved in 10 mM sodium phosphate at pH 6.2 in a 10% 2H2O/90% H2O mixture to 0.5–1 mM final protein concentration. All NMR experiments were performed on a Bruker 600 MHz spectrometer at 310 K. NMR spectra were referenced to external 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). NMR data processing and analysis were performed with the X-PLOR software package (Bruker) and NMRPipe (Delano Scientific). The distance restraints were calculated using the Rosetta software package (Briones et al., 2008). The solution structures of PSV VPg were determined using the program Coot (Emsley et al., 2010).

![Fig. 2. Solution structure of PSV VPg. (A) Stereopair of backbone heavy atoms (N, Cx, and C') for the 20 converged line structures of PSV VPg. N and C indicate N- and C-terminal positions, respectively. (B) Lowest-energy ribbon structure of PSV VPg. N and C indicate N- and C-terminal positions, respectively. α1, α2, and α3 indicate helix 1, helix 2, and helix 3, respectively. The side chain of Tyr956 is shown as a red stick. (C) Hydrophobic patch on the surface of PSV VPg. Amino acids on the surface are colored according to type: magenta, nucleotide acceptor Tyr; yellow, hydrophobic residues (Met, Leu, Ile, Phe, Trp, Tyr, and Val); blue, basic residues (Arg and Lys); red, acidic residues (Asp and Glu); white, other residues. The surface hydrophobic patch is composed of Leu951, Tyr956, Trp959, and Val972 and surrounded by a dashed white line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)](image-url)
analysis were performed using NMRPipe [25] and NMRView [26]. Backbone resonance assignments were obtained using standard triple-resonance NMR experiments as described previously [27]. Assignments of side chain resonances were made using 3D HCCH–COSY and HCCH-TOCSY experiments [28]. Distance restraints were derived from 3D 15N-edited NOESY-HSQC ($t_m = 150$ ms) and 13C-edited NOESY-HSQC ($t_m = 150$ ms) spectra.

2.3. Structure calculations

NOE resonance assignments and initial NOE constraints were obtained from CYANA 2.1 [29] with CANDID [30]. All NOE constraints were manually confirmed during CYANA calculations. Hydrogen bonds for the helical regions of VPg were added to facilitate the automated assignment of additional NOE constraints by CANDID. Chemical shift-based restraints from TALOS$^+$ [31] were included only for secondary structural regions. The initial 100 structures were calculated by a simulated annealing protocol with Xplor-NIH [32]. Iterative refinement and editing of the distance constraints based on the NOESY spectra to remove incorrect and ambiguous assignments reduced the number of constraints. The final 20 structures with the lowest energy were chosen for analysis and were deposited in the Protein Data Bank (accession code 2MXD). All images were prepared with MOLMOL [33].

3. Results

3.1. Construction of the stable core domain of PSV VPg

To determine the PSV VPg structure, we initially cloned full-length VPg (residues Ala935–Glu1048) (Fig. 1A) from the PSV genome and expressed the protein in E. coli. However, the recombinant full-length VPg protein was quickly degraded. Therefore, we analyzed the degraded products using SDS-PAGE and mass spectrometry to identify a stable shorter domain and prepared several shorter constructs (Fig. 1A). Of these constructs, the NMR peaks of 1H–15N HSQC of Ala948–Ser1006 showed great dispersion (Fig. 1B), indicating that this short VPg construct has a compact, well-folded structure in the NMR buffer. Hence, we performed the structural study of PSV VPg using this construct (hereafter referred to as the VPg core domain, VPgC).

3.2. Structure determination of PSV VPgC

The NMR structure of VPgC (Ala948–Ser1006) including three additional amino acids (Gly-Ala-Met) from cloning at the N-terminus was determined using constraints derived from heteronuclear multi-dimensional NMR experiments with uniformly 15N- or 15N- and 13C-labeled VPgC. We obtained assignments for the nearly complete backbone ($\alpha_1$, $\alpha_2$, $\alpha_3$, $\alpha_4$) and ~90% of the side chain assignable protons. The structure was calculated based on the NMR data and refined using CYANA 2.1 and CANDID. The final structure was selected based on the NOE-ROE distance restraints and is shown in Fig. 3A.

**Fig. 3.** Comparison of VPg structures. Conserved hydrophobic core residues of PSV (A), FCV (PDB code: 2M4H) (B), and MNV (PDB code: 2M4G) (C) VPgs are shown as sticks. Helices are indicated as $\alpha_1$, $\alpha_2$, and $\alpha_3$. The nucleotide-accepting Tyr residues are labeled in magenta. The salt bridge between Arg52 and Asp48 in MNV VPg is indicated with a black dotted line. (D) Superposition of the structured cores of PSV VPg (gray) and FCV VPg (gold). (E) Superposition of the structured cores of PSV VPg (gray) and MNV VPg (light blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
on $^1$H–$^1$H NOE constraints. The final ensemble of 20 structures shows excellent structural statistics and structural convergence (Fig. 2A) with final pairwise backbone and all heavy atom root-mean-square (RMS) deviations of $0.55 \pm 0.16$ Å and $1.24 \pm 0.17$ Å, respectively (Supplementary Table 1). The residues in the disallowed region of the Ramachandran plot are located in the disordered N- and C-termini of the refined PSV VPg structure.

### 3.3. Structure description of PSV VPg$c$

The PSV VPg$c$ structure is composed of three $\alpha$-helices ($\alpha_1$, $\alpha_2$, and $\alpha_3$) comprising Asp953–Trp966, Val972–Leu983, and Pro989–Lys1003, respectively (Fig. 2B). The three $\alpha$-helices were tightly packed through hydrophobic interactions formed by seven non-polar residues (Ile962, Met970, Phe975, Leu978, Tyr994, and Trp997). All hydrophobic residues in the structural region of PSV VPg$c$ are buried inside the molecule except for Tyr956 that is positioned in the beginning of $\alpha_1$ (Fig. 3A). The aromatic ring of Tyr956 is unusually exposed to solvent and occupies a small hydrophobic patch on the surface with Leu951, Trp959, and Val972 (Fig. 2C). Negatively and positively charged residues are located around Tyr956. A more detailed discussion of the surface properties of PSV VPg$c$ is included in the Discussion section.

### 4. Discussion

The VPgs of vertebrate viruses including picornavirus and calicivirus are distinct with respect to size, sequence, and structure. The recently reported FCV and MNV VPg structures [24] show a compact helical core flanked by disordered N- and C-termini. The core of FCV VPg comprises a well-defined three-helical bundle, whereas the core of MNV VPg is composed of two helices that structurally correspond to the first two helices of the FCV VPg (Fig. 3B and C). Despite their structural differences, hydrophobic residues in helices 1 and 2 of FCV VPg are relatively conserved in MNV. The two-helix core in the MNV structure is further stabilized by a salt bridge between Arg32 in helix 1 and Asp48 in helix 2 (Fig. 3B). Since the N- and C-termini of the full-length PSV VPg protein were unstable and easily degraded, probably due to their unstructured characteristics, only the stable core region (VPg$c$, residues Ala948–Ser1006) was used for structure determination in this study. The overall conformation of PSV VPg$c$ is similar to that of FCV VPg$c$, which has a compact three-helical core structure stabilized by hydrophobic interactions. The hydrophobic residues that form the compact cores of PSV VPg and FCV VPg are largely conserved: residues Ile962, Met970, Phe975, Leu978, Ala982, Tyr994, and Trp997 of PSV VPg overlaid closely with His30, Leu38, Phe43, Leu46, Phe62, and Trp65 of FCV VPg, respectively (Fig. 3A, b and D). The primary sequences of calicivirus VPg proteins contain a conserved D(E/D)YDEφ sequence motif around the proposed RNA linkage site, where φ represents any aromatic residue. PSV VPg also includes this conserved sequence motif (DDEYEφ), which is identical to that of FCV VPg (Fig. 4A). Notably, the Tyr residue of VPg in this conserved motif functions as a nucleotide acceptor for viral replication and translation [15–19]. With respect to tertiary structure, the Tyr956 within this conserved motif in PSV VPg$c$ is fully exposed to solvent and forms a conserved surface patch at the same spatial position as those of Tyr24 in FCV and Tyr26 in MNV. Although the nucleotidylation site of MNV VPg is controversial [16,19], these conserved sequential and structural features strongly suggest that the exposed PSV VPg Tyr likely facilitates nucleotidylation by viral RNA polymerases.

Structural studies of RNA-dependent RNA polymerases (RdRps) propose that the enzymatic mechanism of nucleotidyl transfer is largely conserved [34]. MNV RdRpo contains conserved YGDD and DYTRWD sequences within the palm domain that mediates catalysis of the phosphoryl transfer reaction. These conserved

---

**Fig. 4.** Comparison of the amino acid sequences and the surface profiles of the structured cores of PSV, FCV, and MNV VPgs. (A) Amino acid sequence alignment of PSV, FCV, and MNV VPgs. (B) PSV VPgC, (C) FCV VPgC, and (D) MNV VPgC domains. The sequences were aligned using ClustalW. The conserved motifs [D(E/D)EYDEφ] are underlined. Conserved amino acids of the structured cores and surface are colored according to type: yellow, hydrophobic residues (Leu, Ile, Val, Met, Tyr, Trp, and Phe); red, negatively charged residues (Asp and Glu); blue, positively charged residues (Arg and Lys). The nucleotide-accepting Tyr residues are shown in magenta. Alpha-helical regions of VPgs are shown over the alignment (PSV and FCV) and under the alignment (MNV). The conserved hydrophobic core residues shown in Fig. 3 are indicated by hyphens above the sequence. (B), (C), and (D) are surface profiles of PSV, FCV, and MNV VPgs, respectively. Surface amino acids are colored according to Fig. 2C. Conserved surface residues are labeled. (E) Representative diagram of conserved surface profile among PSV, FCV, and MNV VPgs: (+) for acidic, (−) for basic, (Ω) for hydrophobic surface, and Y for nucleotide acceptor Tyr residue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
sequences are also found in the closely related HuNV, RHDV, and SaV RdRps [24,35]. In particular, the aspartate residues within these conserved sequences play critical roles in RNA synthesis of (+) RNA viruses. VpGs are nucleotidyalted by viral RdRp, suggesting that RdRps directly recognize the nucleotidylation site on VpGs. The surfaces around the nucleotidylation sites of PSV, FCV, and MNV VpGs share similar structural features (Fig. 4B–D), with the central nucleotide acceptor Tyr residues on the surface surrounded by negatively charged, positively charged, and hydrophobic residues in a counter-clockwise manner (Fig. 4E). The involved surface residues are commonly located in the first and second helices (Fig. 4A), indicating that the conserved surface properties are important for the biological activity of calicivirus VpGs. Although structural information for the calicivirus RdRp–VpG complex is not available, it is likely that the conserved RdRp active site directly interacts with the conserved surface of the VpGs.

Studies on calicivirus and norovirus VpG proteins suggested that they may recruit the ribosome to the genomic RNA by interacting with translation IFs such as the cap-binding protein eIF4E [20,21]. Furthermore, MNV VpG interacts directly with eIF4G without affecting the VpG-eIF4E interaction [13]. Using yeast two-hybrid experiments, a complex between HuNV VpG and the eIF3d component of the eIF3 complex has also been identified [36]. Together, these studies suggest that the VpG protein interacts with a number of binding partners including RNA replication and protein synthesis machinery, forming a hub for assembly of replication and translation complexes [13]. We recently demonstrated that the interaction between PSV VpG and eIF4E was necessary for PSV translation initiation [22]. Moreover, Chung et al. recently reported that the unstructured C-terminal tail of MNV VpG binds directly to the eIF4F complex through interactions with eIF4G [23]. Although the functional roles of the N- and C-terminal flexible regions of PSV VpG have not yet been confirmed, it is likely that they may interact with various binding partners (such as eIF4E) of VpG proteins. In conclusion, the solution structure of PSV VpG reveals conserved sequential and structural features among calicivirus VpGs and suggests a conserved mechanism of replication and translation in this viral family.

Conflicts of interest

The authors disclose no potential conflicts of interest.

Acknowledgments

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology of Korea (NRF-2013R1A1A2009419 to C.W.L and 2014R1A2A201004915 to J.-S. K.) and the Ministry of Science, ICT and Future Planning (NRF-2014M2B2A4029302 to C.W.L.).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jbbrc.2015.02.156.

Appendix B. Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.jbbrc.2015.02.156.

References


